Abstract

This paper concerns the design and analysis of a ball type continuously variable transmission (B-CVT). This B-CVT has a simple kinematic structure, and same as a toroidal CVT, transmits power by friction on the contact points between input and output discs, that are connected to each other by balls. After, a brief introduction of our B-CVT structure, the performance and traction efficiency of B-CVT is analyzed. The geometry and speed ratio of the proposed CVT is obtained. Then, by finding the contact areas between rotating elements and stress distribution through them, the torque capacity of B-CVT is computed. Next, the power loss of the system caused by various parameters such as relative arrangement of rotating elements as well as relative velocity at contact areas is found. Finally, after presenting the influence of the different geometrical and assembly conditions at efficiency of the system, the efficiency of the system compared with the efficiency of a Toroidal CVT.

Keywords: Ball CVT, Traction, Efficiency, Geometry

1. Introduction

Nowadays, CVT systems have found their place in automotive and other industries. Currently, besides manual and automatic transmission systems, some car makers use continuously variable transmission systems in their products. This is due to simpler and softer performance as well as higher efficiency of the CVTs. Higher efficiency in CVTs decreases fuel consumption up to 10% and reduces greenhouse gases propagation to the atmosphere [1]. Commercially, there are two kinds of CVTs, named as belt CVTs and toroidal CVTs. The first kind consists of a pair of V-shaped pulleys which are connected to each other by a plastic or metal belt. One half of these pulleys are able to move axially in order to create different speed ratios, and power is transmitted from input pulley to the output one through the belt by frictional contact with pulleys. Currently, this kind of CVT is employed by the vehicle companies such as Honda, Ford, and Toyota as well as small trucks. There are some researches who investigate the power loss mechanisms due to friction within the belt drive CVTs [2,3]. Meanwhile, toroidal CVTs are made up of disks and rollers that transmit power between the disks. One disk is the input, and the other is the output. Power is transmitted from input disk to the output disk by the rollers, through a limited number of lubricated contact areas. Some companies same as Nissan, NSK and Torotrack have developed different kinds of toroidal CVTs for vehicle and other industries.

The lubrication of contact areas is under very severe stress and thermal conditions with contact pressure near to 1-3 Gpa [4]. Because of the existence of thin layer of lubrication at contact areas, we may see a little power loss due to the slippage. Spin loss is another source of energy loss in toroidal CVTs. Carbone et al. analyzed the efficiency and performance of half toroidal CVT and compared it with the full toroidal one. They found that half toroidal CVT shows better performance and higher efficiency than full toroidal one [5]. L. De novellis et al. investigated the efficiency and traction capability of double full toroidal CVT and compared it with both full and half toroidal CVT. They proved that although double full toroidal CVT has zero spin in its neutral position, its efficiency is similar to full toroidal CVT and lower than half toroidal one [4]. They proved that although double full toroidal CVT...
has zero spin in its neutral position, its efficiency is similar to full toroidal CVT and lower than half toroidal one. Delkhosh and Foumani introduced an optimization algorithm for high power transmission efficiency, based on the toroidal CVT’s geometry and kinematics [6].

Recently some attentions have been paid to another type of traction drive transmission systems named as Ball CVT (B-CVT). The B-CVT is intended to overcome some of the limitations of existing CVT designs. Its compact and simple design and relatively its easier control make a good potential that B-CVT to be used in wide variety of mechanical and vehicle transmission systems. Working principle of this system is like a toroidal CVT. Basically, main components of B-CVTs consist of input disk, output disk and balls. B-CVT is traction type and balls work instead of rollers in toroidal CVTs, which connect these two disks to each other. Pohl et al. presented the analogy between the B-CVT and a controversial planetary gear set. [7]. Carter et al. has analyzed the effect of B-CVT usage on performance and efficiency of a two-Wheeled light electric vehicle. They showed that the B-CVT not only raised the top speed and time driving, but improved the controllability of the vehicle[8]. Park et.al. developed a prototype of a ball CVT for a motorcycle. They determined the design parameters, and measured the efficiency performances of the CVT experimentally [9]. Kim, et al., employed a ball CVT to drive a nonholonomic wheeled mobile robot. Their design had a low power to weight ratio, that makes it unsuitable for automotive or similar heavy duty applications [10].

Although in the recent years the researchers have investigated and designed different kinds of B-CVTs, there is not detailed and analytical description on their traction performance, power efficiency and corresponding power losses. In this paper, we analyze the performance of a B-CVT. Here, the Hertz theory is used to model the pressure distribution over the contact area to compute the power rate and friction losses. The relative velocities at contact areas and related spin loss as main sources of power loss in CVT systems are computed. After describing geometrical and traction parameters we derive power transmission efficiency of the B-CVTs. Finally, the effect of different geometric and power transmission conditions in efficiency of a B-CVT is presented, and compared with a toroidal CVT.

2. Principle and structure of the prototype CVT

Although, in recent years, various configurations of ball CVTs have been introduced, all of these systems have the same working principles. The difference of B-CVTs backs to the diversity in control scheme of ball tilting angle in order to achieve the different speed ratios. The operation of a B-CVT has been depicted in Figure 1. Speed ratio is controlled by tilting the balls rotation axis angle $\gamma$ which leads to 3 different conditions. In Figure 1-a, the system is in under drive condition, means the output shaft rotates slower than input shaft. In Figure 1-b, the system is in neutral position, so the speed of the input and output shafts are the same, and in Figure 1-c the system speed ratio is upper than 1.

In B-CVTs, the contact angle $\alpha$ between ball and disk contact line and horizontal line, is another geometric parameter that has direct effect on angular velocity of the ball. The effect of this angle on B-CVT performance will be discussed in detail in the next sections. In this kind of CVT, the balls are not translated but rotate about axes passed through their centers. B-CVTs usually includes 4 balls to transmit power between input and output disks, but CVTs with 6 and 8 balls are used for specific applications.

![Fig1. Different situations of ball CVT](image-url)
As shown in the Figure 2, we made a B-CVT prototype consists of rotating input and output discs connected to the input and output shafts, respectively, while the balls transmit power between them. Each ball assembled on a shaft and can rotates about it. The transmission ratio is controlled by tilting the angle of balls via a slotted control plate rotated by a worm gear and is able to tilt ±30° to generate desired speed ratio equal to 5 between input and output shafts. In the Figure 2-b an exploded view shows different parts of the B-CVT as A) output section consists of control system, B) Power transmission section consist of disks and balls, and C) input section. A ball bearing encloses the balls to limit balls position and to obtain a surface to make necessary normal force at contact point between balls and disks. A preload to obtain traction force between balls and disks is applied by compression of a helical spring (not seen in the Figure 2) with a nut.

3. Geometric description

As it is mentioned earlier, in both toroidal and B-CVTs, power is transmitted by the traction (See Figure 3). The geometry of a ball CVT is characterized by radii of curvature of ball and disks. The first principle radii of the system are the input and output radii of the disks $r_1$ and $r_3$ respectively, as well as the radius of the disk contact point radius of curvature $r_0$ and the radius of the ball $r_b$. Whereas $r_{21}$ and $r_{23}$ are the normal distances between the rotating axes of the balls and input and output disks respectively (see Figure 3-b). The conformity ratio factor $CR$ of the system is defined by $CR = r_b / r_0$. The geometric data for this B-CVT is given in the Table 1.

Table 1. Geometric Data of B-CVT Prototype

<table>
<thead>
<tr>
<th>Disks mean radii (mm)</th>
<th>Ball radius (mm)</th>
<th>Angle $\alpha$ (deg)</th>
<th>Limits of Balls tilting angle (deg)</th>
<th>disks curvature radius (mm)</th>
<th>Conformity ratio $CR = r_b / r_0$</th>
<th>No. of balls</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>25</td>
<td>60</td>
<td>±30</td>
<td>40</td>
<td>0.625</td>
<td>4</td>
</tr>
</tbody>
</table>

(a)
By tilting the ball, dimensions of $r_{21}$ and $r_{23}$ changes

$$r_{21} = r_b \sin(y + \alpha)$$  \hspace{1cm} (1)

$$r_{23} = r_b \sin(\alpha - y)$$  \hspace{1cm} (2)

Considering pure rotation in contact areas, the ideal speed ratio (SR) of the system is defined as:

$$SR = \frac{\omega_3}{\omega_1} = \frac{\sin(\alpha - y)}{\sin(\alpha + y)}$$  \hspace{1cm} (3)

4. Contact model

In order to evaluate the performance of a ball CVT, we need to calculate the contact pressure at the interface between the balls and disks. Then, we can determine tangential force to transmit power, as well as spin loss. At next stage based on these computations the rated power of B-CVT as well as efficiency could be evaluated. As shown in Figure 4,
a suitable coordinate system on contact surface should be defined. In this coordinate system the x axis lays in the moving direction and the z axis is perpendicular to the contact surface. We define reduced radii along the x and y axes

\[
\frac{1}{r_{eqx}} = \frac{1}{r_{ax}} + \frac{1}{r_{by}}
\]

(4)

\[
\frac{1}{r_{eqy}} = \frac{1}{r_{ay}} + \frac{1}{r_{by}}
\]

(5)

Where the subscript a refers to the disk and b to the ball, and the radii of the curvatures are determined from Figure 9 as, \(r_{ax} = r_{by} = r_{b}\), \(r_{ay} = -r_{0}\) and \(r_{ax} = r_{i}/\sin(\alpha)\). The equivalent radius of the curvature is then defined as

\[
\frac{1}{r_{eq}} = \frac{1}{r_{eqx}} + \frac{1}{r_{eqy}}
\]

(6)

When two elastic bodies with different radii of curvature along the co-ordinate axes pressed together a contact area is generated at the interface of two bodies. The geometry of this surface depends on the normal force \(F\), modulus of elasticity \(E\), and the geometry of two connected bodies. According to researches accomplished by Hamrock and Dowson [11], the semi-axes of the elliptical contact areas can be calculated by

\[
A = 2 \left(\frac{6k^2E_{eq}}{\pi E'}\right)^{1/3}
\]

(7)

\[
B = 2 \left(\frac{6E_{eq}}{E'}\right)^{1/3}
\]

(8)

where

\[
E' = 2 \left(\frac{1-v_a^2}{E_a} + \frac{1-v_b^2}{E_b}\right)^{-1}
\]

(9)

Here, \(E'\) is the modified modulus of elasticity. Also, elliptical integrals of the first \(\epsilon\) and second kinds are

\[
\epsilon = 1 + \left(\frac{z}{a}\right) \ln a
\]

where \(a\) is the ratio of equivalent radii \(a = R_{eqy}/R_{eqx}\) and \(F\) is normal applied load [11].

According to Hertz theory, pressure distribution over contact area which shows the amount of normal stress on any arbitrary point \(\sigma(x,y)\) on contact area is expressed by

\[
\sigma(x,y) = \sigma_{max}\left[1 - \left(\frac{x}{A}\right)^2 - \left(\frac{y}{B}\right)^2\right]^{1/2}
\]

(11)

Where \(\sigma_{max} = 3F/2\pi AB\) is maximum pressure.

5. Spin loss

Since the rotation axes of rotating elements generally are not parallel with each other, an unwanted energy loss generated named as spin loss. Spin is a project of the relative angular velocity between a disk and a ball on direction normal to the contact area. This unwanted velocity causes spin loss which should be reduced as much as possible. Although the existence of the spin is inevitable, we can reduce and even for some speed ratios eliminate the effect of this parameter by designing suitable configuration.

The relative angular velocity vector between input disk and ball \(\omega_{21}\) is shown in the Figure 5. Also, relative angular velocity between balls and output disk \(\omega_{32}\) is \(\omega_3 - \omega_2\). has a similar vector diagram. These two vectors consist of two non-zero components \(\omega_{21\text{spin}}\) and \(\omega_{32\text{spin}}\)

\[
\omega_2 = \omega_1 + \omega_{21\text{spin}} + \omega_{21\text{roll}}
\]

(12)

\[
\omega_3 = \omega_2 + \omega_{32\text{spin}} + \omega_{32\text{roll}}
\]

(13)

Where

\[
\omega_{21\text{spin}} = \omega_1 \cos \alpha + \omega_2 \cos(\alpha + \gamma)
\]

(14)

\[
\omega_{32\text{spin}} = \omega_3 \cos \alpha + \omega_2 \cos(\alpha - \gamma)
\]

(15)

![Fig4. Angular velocity diagram between input disk and balls](image-url)
We define the spin coefficient as a function of the \( \alpha \) and tilting angle \( \gamma \) as

\[
\sigma_{21} = \frac{\omega_{21\text{spin}}}{\omega_{21}} = \cos \alpha + \frac{r_1}{r_2 \tan (\alpha + \gamma)}
\]

\[
\sigma_{32} = \frac{\omega_{32\text{spin}}}{\omega_{32}} = \cos \alpha + \frac{r_3}{r_2 \tan (\alpha - \gamma)}
\]  

In order to provide the maximum speed ratio of this system equal to 4, the angles \( \alpha \) and \( \gamma \) should be selected properly. Beside the speed ratio, this selection must give the lower angular velocity for the balls, because the lower ball angular velocities, results lower spin losses. As mentioned earlier, angle \( \alpha \) has direct effect on angular velocity of the ball. As it is seen in Fig.4, the larger the angle \( \alpha \), the lower the ball angular velocity. According to Eqn. (3), if the ball tilts \( \pm 30 \) deg, the optimal value for \( \alpha \) would be close to 60 deg to provide speed ratio of 4. In this angle the system will provide desirable speed ratio with maximum efficiency. From Eqns. (16) and (17), it can be found that for selected values of angles \( \alpha \) and \( \gamma \), spin components of the relative angular velocities \( \omega_{21\text{spin}} \) and \( \omega_{32\text{spin}} \) never vanishes. Since the angle \( \gamma \) is always smaller than the angle \( \alpha \), the sign of spin is always positive. Figure 6 shows the behavior of the angular velocity of spin at inner and outer contacts and the angular velocity of ball (\( \omega_{21\text{spin}}, \omega_{32\text{spin}} \)) in a ideal speed ratio \( SR = 2 \), for different angles \( \alpha \).

Figure 6. The input and output spins and ball angular velocities as a function of angle \( \alpha \).

Also, the spin coefficients change as Figure 7. It can be observed that because of the symmetry in B-CVT, the spin coefficients behavior are the same but opposite. As it is expected, in neutral position \( SR = 1 \), two coefficients have the same value and they never meet zero.

As mentioned earlier, Equation (11) gives normal stress of any arbitrary point of the contact area. The tangential force is obtained with friction equation \( F_T = f_f F_N \). Since the direction of tangential force is along with relative slippage velocity. Between many experimental velocity dependent friction coefficient models, including the Stribeck effect, the following is the most accepted:

\[
f_f(V) = f_s + (f_s - f_k) e^{-\left(V/V_{str}\right)^2} \
\]

Where, \( f_s \) and \( f_k \) denote static and kinetic friction, respectively, and \( V_{str} \) is the critical Stribeck velocity. Hence, the differential tangential force in any arbitrary point in input and output contact areas is calculated by

\[
dF_{T_{in}} = f_f(x,y) \frac{\omega_{21}}{|\omega_{21}|} dA_{in} \]

\[
dF_{T_{out}} = f_f(x,y) \frac{\omega_{32}}{|\omega_{32}|} dA_{out} \]

Since the relative velocity in contact area is defined by \( v_{rel} = \omega_{rel} \times p(x,y) \), so the relative velocities of the system as a function of spin angular velocities are

\[
\frac{\omega_{21}}{\omega_{21\text{spin}}} = -\gamma \omega_{21\text{spin}}, \frac{\omega_{21}}{\omega_{32\text{spin}}} = x \omega_{32\text{spin}}
\]

\[
\frac{\omega_{32}}{\omega_{32\text{spin}}} = -\gamma \omega_{32\text{spin}}, \frac{\omega_{32}}{\omega_{21\text{spin}}} = x \omega_{21\text{spin}}
\]

and simply

\[
|v_{21}| = \left( v_{21x}^2 + v_{21y}^2 \right)^{1/2} \text{ and } |v_{32}| = \left( v_{32x}^2 + v_{32y}^2 \right)^{1/2}.
\]
By integrating Eqns. (19) and (20), we can obtain the tangential force on overall contact areas

\[
F_{T_{in}} = \mu_{in} \sigma_{max, in} \oint \left[ 1 - \left( \frac{x}{A_{in}} \right)^2 - \left( \frac{y}{A_{in}} \right)^2 \right] \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}} dA_{in}
\]

(23)

\[
F_{T_{out}} = \mu_{out} \sigma_{max, out} \oint \left[ 1 - \left( \frac{x}{A_{out}} \right)^2 - \left( \frac{y}{A_{out}} \right)^2 \right] \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}} dA_{out}
\]

(24)

Where, the subscript \( j = x \) or \( y \). According to Figure 8, spin momentum is calculated by multiplying the tangential force to the distance of \( p(x,y) \) to the center of the contact area, means

\[
dT_{s_{in}} = \mu_{in} \sigma_{max, in} \left[ 1 - \left( \frac{x}{A_{in}} \right)^2 - \left( \frac{y}{A_{in}} \right)^2 \right] \sqrt{x^2 + y^2} dA_{in}
\]

(25)

\[
dT_{s_{out}} = \mu_{out} \sigma_{max, out} \left[ 1 - \left( \frac{x}{A_{out}} \right)^2 - \left( \frac{y}{A_{out}} \right)^2 \right] \sqrt{x^2 + y^2} dA_{out}
\]

(26)

**Performance analysis**

According to the Figure 9, the power is transmitted by tangential force \( F_T \). In the contact areas the spin force as well as the tangential force is generated. The spin forces generate spin moments \( T_{s_{in}} \) and \( T_{s_{out}} \) that are the main reason of power loss in contact areas. Ignoring inertial forces and according to the Fig.8, these forces obtained by writing momentum equilibrium on ball, input and output disks as below

\[
F_{T_{out}} r_2 - F_{T_{in}} r_1 - T_{s_{in}} \cos(\alpha + \gamma) + T_{s_{out}} \cos(\alpha - \gamma) = 0
\]

(26)

\[
T_{in} = n(F_{T_{in}} r_1 + T_{s_{in}} \cos \alpha)
\]

(27)

\[
T_{out} = n(F_{T_{out}} r_2 - T_{s_{out}} \cos \alpha)
\]

(28)
In Eqns. (26-28), \( n \) is the number of balls.

The ideal input power and frictional power loss at the ball-disk contacts, where slip takes place, are

\[
\text{Pin} = nF_{\text{in}}r_1w_{21 \text{roll}} \quad \text{Ploss} = n(T_{\text{pin}} \cos \alpha w_{21 \text{spin}} + T_{\text{out}} \cos \alpha w_{22 \text{spin}})
\]

Therefore, mechanical efficiency of B-CVT is calculated as

\[
\eta = \frac{\text{Pin} - \text{Ploss}}{\text{Pin}}
\]

Using the procedure described above, the performance of a B-CVT has been evaluated. Here, we analyze the effect of geometrical parameters same as ball tilting angle \( \gamma \), relative contact angle \( \alpha \), and conformity ratio in the B-CVT efficiency. Input angular velocity \( \omega_1 \) is considered to be constant and equal to 3000 rpm. Meanwhile, the geometric parameters of the system presented in table 1.

The contact area increases by applied normal load, correspondingly, which leads to more spin loss. Here, material of disks and balls are selected to be heat treated 4041 steel, so the normal allowable load is chosen such that maximum allowable stress in contact areas become less than the material yield stress.

5.2 The Effect of Tilting and Contact Angles

Fig. 9 shows efficiency in terms of tilting angle \( \gamma \) for different contact angle \( \alpha \). It is observed that increasing relative contact angle between balls and disks \( \alpha \) will improve the overall efficiency of the system. Because, increasing the \( \alpha \) angle decreases the ball angular velocity, and according to Eqns. (14) and (15) has direct effect on the spin angular velocities. This is clear that, the higher spin angular velocities result in the lower system’s efficiency because of higher spin losses. As it is seen in Figure 10, for a specific input torque and in specific contact angle \( \alpha = 60^\circ \) and given tilting angle \( \gamma \), the power loss decreases with increasing tilting angle \( \gamma \).
Fig 9. Efficiency in terms of $\gamma$ for different $\alpha$ angle

Fig 10. Efficiency in terms of tilting angle $\gamma$ for three different input power.

Fig 11. The effect of conformity ratio CR on efficiency.
5.3 The effect of conformity ratio

Conformity ratio \( CR = \frac{r_0}{r_2} \) is another geometrical parameter with significant effect in the systems efficiency. The larger value of \( CR \) correspond with larger contact surface between disks and balls, that results more power lose and lower efficiency (Figure 12). Conversely, for desired power transmission, smaller \( CR \) values consequences a smaller contact surface that results higher normal stress than may be more than allowable stress at contact area. Since selecting heat treated 4340 steel as balls and disks material, hence \( CR = 0.625 \) was find out at the contact areas encounter allowable stress.

5.3 Comparing B-CVT with toroid CVT efficiency

Here, to find the frictional losses and efficiency of toroidal CVT, we utilized the same method explained in sections 4 and 5 for B-CVT. According to the Figure 2, the geometric properties of the full toroidal CVT presented in Table 2, is comparable with our made B-CVT prototype represented in Table 1. Hence, the values of the curvatures and the input power of the both systems have been selected to match with each other.

We have found that because of the special geometry of toroidal CVT, for the same power transmission ratio, the normal load at contact area of full toroidal CVT is 2 or 3 times more than the B-CVT’s. This results the higher spin loss in full toroidal CVT. Such that, spin coefficients in toroidal CVTs are less than B-CVT. Our computations show that, the spin coefficients, in toroidal CVT contact areas, vary between ranges 0.86-1, which according to Figure 5, and are less than B-CVT’s.

Figure 13 depicts the efficiency of the two systems when both CVTs are operated with their maximum power.

As can be seen in Figure 13, in the worst circumstance, the difference between the efficiency of two systems is near 3.6%. Also, in the full operating range of tilting angle \( y \) the average efficiency of two systems are approximately equal to 98.1%.

<table>
<thead>
<tr>
<th>No. of Rollers (N)</th>
<th>Ball tilting angle (deg)</th>
<th>Conformity ratio ( CR = \frac{r_2}{r_0} )</th>
<th>Aspect ratio ( k = \frac{e}{r_0} )</th>
<th>Roller radius ( r_2 )(mm)</th>
<th>disks cavity radius ( r_0 )(mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>±30</td>
<td>0.625</td>
<td>0.25</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2 geometry properties of Full Toroidal CVT
6. Conclusion

In this paper, we introduced a new type of a ball CVT and developed a basic methodology to analyze the performance and efficiency of the ball CVT. Our analysis shows that relative geometrical dimension and arrangement between input and output disks with balls has significant effect on the overall efficiency of the ball CVT. Also, comparing a ball CVT with a full toroidal CVT under equal conditions shows that efficiency of both systems are similar. Since the geometry of different kinds of ball CVTs are such that they are easily controllable, the B-CVT has a good potential to be used instead of belt and toroidal CVTs.

References